DOCUMENT RESUME

ED 173 363 TM 008 529

AUTHOR Becker, Wesley C.
TITLE Symposium on Direct Instruction: The Follow Through Data Show That Some Programs Work Better Than Others.
PUB DATE Mar 78
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); For Abt Report, see CD 148 490-493; For House et. al.'s critique, see EJ 173 526

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Affective Behavior; *Basic Skills; *Compensatory Education Programs; *Data Analysis; Economically Disadvantaged; Educationally Disadvantaged; *Effective Teaching; Models; National Norms; National Programs; Primary Education; *Program Effectiveness; Research Design; *Research Problems; Research Reviews (Publications); Teaching Methods
IDENTIFIERS *Project Follow Through

ABSTRACT

The Abt Associates' final report on the Follow-Through Planned Variation Experiment is defended against the critique published by a blue ribbon panel funded by the Ford Foundation. In contrast to the conclusions drawn by House, Glass, McLean, and Walker that Follow Through showed no "winners" or "losers", the National Evaluation significance data show clearly that direct instruction outperformed other approaches on nearly all measures. Furthermore, a simple examination of normative data for the direct instruction model in comparison to eight other major models, shows that the children in the direct instruction model are at or near national norms on all Metropolitan Achievement Test (MAT) measures at the end of third grade. The conclusions drawn from these analyses are also supported by a reanalysis by Bereiter and Kurland. (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED173363

SYMPOSIUM ON DIRECT INSTRUCTION

The Follow Through Data Show that Some Programs Work Better than Others[1]

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Wesley C. Becker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Wesley C. Becker
University of Oregon
College of Education
Division of Teacher Education

Paper presented at AERA, Toronto, March 1978
All rights reserved - W. C. Becker

TM008 529

The March 1978 issue of the Phi Delta Kappan presents the conclusions drawn by a blue ribbon panel funded by Ford Foundation to critique the Abt Associates' final report on the Follow Through Planned Variation Experiment (Abt Associates, Inc., 1977). The major conclusions of the critique by House, Glass, McLean, and Walker (1977, 1978) are:

1. Contrary to the Abt conclusions, models that emphasize basic skills are "no more effective than others."
2. That "participation in Follow Through classes was not shown to produce results either superior or inferior to those of schooling normally provided" (emphasis added).
3. That they agree with a "major finding" of the Abt study, namely, "The effectiveness of a teaching approach varies greatly from one school district to the next. This finding should be honored widely and should serve as a basis for educational policy" (1978, p. 473; emphasis added).

We totally disagree with these conclusions.

The critique appropriately points out classification problems by Abt Associates in summarizing the data, but these problems have nothing to do with the data itself. The critique compounds the Abt errors by adding its own brand of inappropriate classifications. For example, they label basic skills as those which can only be taught by rote methods, and assert that "basic skills models" use only rote methods.

The critique's analyses of "measurement" and "fairness" issues turn out to be based on self-contradicting statements and clearly post hoc arguments.

In attacking the main Abt analysis strategy, House, et al. assert the need for a statistical method not influenced by sample size. They contend that the Abt method selectively favored "basic skills" models. However, sample size has nothing to do with direction of effects. The embarrassing problem for several models currently popular in education is their strong negative effects, not the null effects. Sample size does not operate selectively on positive effects only. Technically, the solution "offered" to solve Abt's problems is inadequate for at least four reasons: (1) the preschool adjustments of the data are not used, (2) Cohort II Entering-K data are ignored arbitrarily, (3) the analysis confounds sponsor with duration of treatment (since the number of K-starting and first-starting sites varied greatly by sponsor), and (4) the method fails to adjust for cases where the FT-NFT groups are so mismatched that no statistic could undo the damage. The revised analysis shows little understanding of the basic data set it was designed to cope with, especially the need to carefully consider the adequacy of the matches between Follow Through and Non-Follow Through groups in each local analysis.

The House conclusion that "nothing was demonstrated" is best refuted by examining the Abt findings presented for the nine major sponsors (for whom there were at least eight control-group comparisons in the National Evaluation).

The major findings of the Abt Report are given in a series of tables, one for each sponsor, as illustrated in Table 1. For each variable, a covariance adjusted comparison was made with a local comparison group

Insert Table 1 about here

TABLE 1
SUMMARY OF EFFECTS
FOR DIRECT INSTRUCTION SITES

Domain	Outcome Measure	Stream	KINDERGARTEN												ENTERING FIRST											
		Site	New York, NY		Grand Rapids, MI		W. Iron Co., MI		Flint, MI		Providence, RI		E. St. Louis, IL		Racine, WI		Dayton, OH		Tupelo, MS		Williamsburg Co., SC		E. St. Louis, IL			
		Cohort	III		II		III		II		III		III		II		II		III		III		II		III		II	
		Analytic Comparison	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled	Local	Pooled
BASIC SKILLS	Word Knowledge			□		-				-		+		⊞		+	-	-			-		-	+	+	+		+
	Spelling			□		-				-		⊞		⊞		-	-	-				+	+	+	+	+		
	Language	+	+	+	+	+		-	+		+	+	+	+	+	+			+	+		♦		+	+	+	+	+	+
	Math Computations	+	+			+					+	+	+	+	+	+			+	+				+	+	+	+	+	+
COGNITIVE CONCEPTUAL SKILLS	Raven's							-			-	-				-	-	-			-	-	-	-				
	Reading					+	⊞			-		+		⊞		+	-	-			-		-		+	+	+	+	
	Math Concepts	⊞	+				□	-	□	-	+	+	+	+	-	+			+	+			-	+	+	+	♦	♦	+
	Math Problem Solving	⊞	+			♦			□	-	+	+	+	+		+			+	+		♦		+	+	+	+	♦	+
AFFECTIVE MEASURES	Coopersmith		-			+		-		+	-	-				+		+		-				+	+	+		+
	IARS (-)			□				+						+											+	+		
	IARS (+)									⊞				⊞					+	+	-	-			+	+		
	Total Reading			□		+				-				⊞		+	-	-			-		-		+	+	+	+	+
	Total Math	+	+			+		-		-	+	+	+	+	+	+			+	+			-	-	+	+	+	+	+
	Language Part A		+			+						⊞	+	+	+	+			+	+		+			+	+	+	+	+
	Language Part B	+	+	+	+	+		-			+	+	+	+	+	+			+	+		+		+	+	+	+	+	+

NATIONAL NORMS*		Site FT Sample*																	
Outcome Measures	MAT Sample																		
Reading	20%	18	10	41	40	12	3	17	25	18	21	43	41	42	45	60	43		
Math	12%	7	10	41	44	0	9	10	14	2	5	13	20	20	33	50	15		
Spelling	22%	18	6	36	33	6	6	20	27	21	23	45	38	41	38	53	31		
Language**	14%	11		77		12	17		22			30	25	22	42	30	15		

Notes

* In terms of percentage of children performing one or more years below grade level.
No national norms comparisons were made for Cohort II-K children on the Language subtest

\+ FT-favoring effect
\- NFT-favoring effect
blank Null effect

□ Preschool Adjustment changes effect
□ Untrustworthy effect

and with a pooled national comparison group. When FT exceeded NFT by at least one-fourth standard deviation on a given variable, and when the difference was statistically significant, this was considered an educationally significant outcome and a plus (+) was placed in the table. When NFT exceeded FT by the same criteria, this was considered to be a significant negative outcome and a minus (-) was placed in the table. When the results were in between these limits, the difference was considered null and the table left blank. At the third grade level, the one-fourth standard deviation criterion corresponds to about two months difference in grade-norm terms.[2]

Table 1 displays the Abt findings for each Direct Instruction site in the National Evaluation. The kindergarten-entering sites are shown first, followed by the first-grade-entering sites. At the bottom of the table, the percent of students one year or more below grade level on major MAT measures is given. In the draft version of this table, median grade equivalents were given for each site. We will use the medians from the draft report to summarize the normative level of performance for each model.

Grayouts. The shaded areas in the tables are termed "grayouts." Effects were grayed out for one of two reasons. First, a whole column was grayed out when the percent of children attending preschool differed for FT and NFT by more than 50%. Second, a specific effect was grayed out when the covariance adjustment changed a site mean by more than one-half standard deviation. The Abt report calls the grayed out effects "untrustworthy." We believe that this is an inappropriate conclusion when a plus (+) FT effect is found under conditions where the operating biases favor a minus (-) effect (or vice versa). For example, the nature

of the correlations between preschool attendance and outcome measures favor NFT. Under such conditions, plus (+) FT findings are actually conservative estimates of real outcomes. This is the case for each of our sites grayed out in Table 1. Furthermore, the magnitude of the actual "preschool effect" is so small (the highest correlation is -.09) that elimination of data for this reason is questionable.[3]

The elimination of all outcomes where the covariance adjustment exceeds one-half standard deviation on the outcome measure can also be questioned. As Campbell and Erlebacher (1970) point out (and the Abt report acknowledges), the effect of a covariance adjustment for the group initially lower on the covariates is to underestimate the correction. "This bias is due to error and uniqueness in the covariate, i.e., variance not shared by the dependent variable" (p. 199).

Insert Table 2

The appendices to the Abt IV Report show the direction and magnitude of the adjustment for each variable. In Table 2, we have shown the relation between direction of adjustment (positive, negative) and outcome effect (+, 0, -). Where an adjustment is positive (FT is initially lower than NFT on covariates) the built-in bias should favor minus (-) outcomes for FT. Where an adjustment is negative (NFT initially higher than FT), the built-in bias should favor plus (+) outcomes for FT. The arrows in Table 1 show the direction of the expected bias. The heavy outline boxes cover effects operating against the direction of the covariance adjustment bias. The data in Table 2 clearly show that for the Direct Instruction outcomes, this grayout rule is inappropriate. A more conservative procedure would not gray out the effects in the heavy outlined

Table 2

Tabulation of Relation of Outcome Effects and

the Direction of Covariance Adjustments for Grayouts

Based on Magnitude of Adjustment

	Direct Instruction			Eight Other Major Sponsors			
Effect	Adjustment Positive	Adjustment Negative	Totals	Effect	Adjustment Positive	Adjustment Negative	Totals
+	35	1	36	+	32	3	35
0	28	1	29	0	96	29	125
-	10	3	13	-	45	24	69
Totals	73	5	78		173	56	229

boxes, and would indicate that the remaining grayed results (on the average) underestimate true FT effects, rather than being called "untrustworthy." The general effect of applying this grayout rule is to attenuate real differences between model sponsors.

To summarize model effects, the authors of the Abt report counted the number of pluses for a given group of measures (such as Basic Skills), subtracted the number of minuses, and divided by the number of comparisons. Both local and pooled comparisons were included. Sites were counted if the number of grayed-out effects was fifty percent or less, but only the non-grayed-out effects were counted. Engelmann and Becker (in press) independently adopted a similar index which they gave the name Index of Significant Outcomes (ISO). It was basically the index developed by Abt multiplied by 1000 to eliminate decimal points.[4] For a given group of effects, if a sponsor is more positive than negative, the ISO is positive. If the average effect is more negative than positive, the ISO is negative. The ISOs permit a relative comparison of programs and of skill areas within programs.

In the tables which follow we will present the ISO values for sponsors on Basic, Cognitive, and Affective Measures as given by these three rule systems:

1. Abt IV rules: (a) eliminate the site if more than 50 percent of effects are grayed out; (b) count only non-grayed-out effects; (c) count both local and pooled comparisons.
2. Use all data: count all pluses and minuses in tables to compute ISOs.
3. Selectively keep grayouts: (a) keep all data grayed out for the preschool difference; (b) keep those pluses and minuses where the effect is in the opposite direction to the expected bias (Table 2, heavy outline boxes); (c) eliminate site if more

than 50 percent of effects are still grayed out; (d) count only non-grayed out effects with a and b above treated as non-grayed out.

Additionally, the figures show the effects for Direct Instruction with and without the inclusion of Grand Rapids.[5] The Cognitive Measures are graphed without the questionable Raven's Coloured Progressive Matrices.

Basic Skills

Figure 1 and Table 3 show the results for the various analyses of relative outcomes. As is apparent from Figure 1, Direct Instruction is the only model to show substantial positive effects. Parent Education, Behavior Analysis, and SEDL show an average effect near zero, and all other models show substantial negative effects (performances below the level of their comparison groups after the statistical adjustments for initial inequalities). Using the Abt IV rules, the ISOs range from +309 for Direct Instruction to -274 for Open Education (EDC). The All Data analysis shows a similar result for Direct Instruction as the Abt IV analysis, but lowers the ISOs for four of the five models with negative effects.

Insert Figure 1 and Table 3

The Selective Keeping of Grayouts analysis improves the ISO for Direct Instruction (+399) and gives middle-ground values to models with largely negative outcomes. When Grand Rapids is eliminated from the Direct Instruction Model data, the ISOs range from +366 to +419 for the various analyses, a substantial improvement.

FIGURE 1

ISO's for Basics Measures

(Word Knowledge, Spelling, Language, Math Computation)

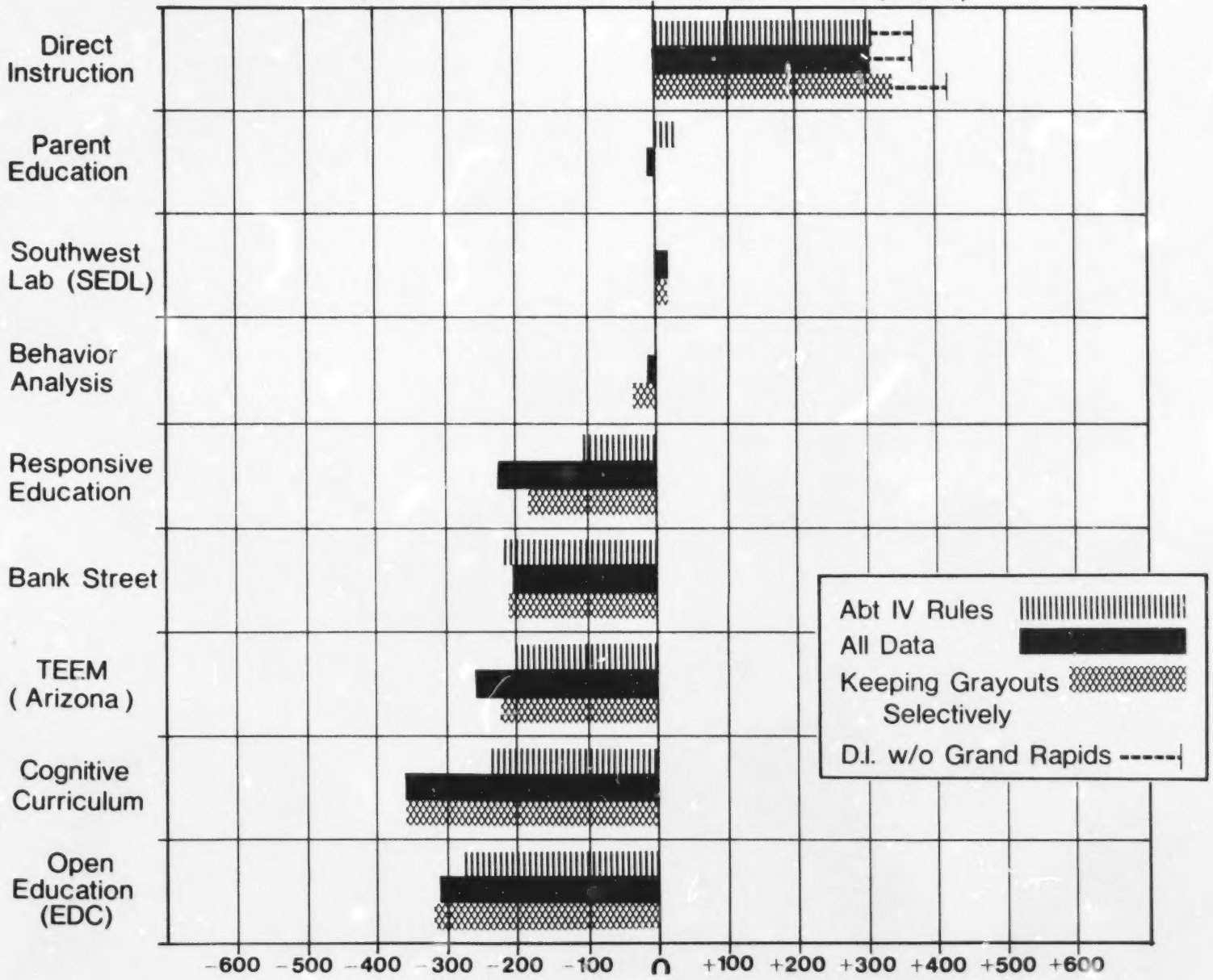


Table 3

ISOs for Basic Skill Measures Using the All Data Analysis

	No. Sites	Word Knowledge ISO	Rank	Spelling ISO	Rank	Language ISO	Rank	Math Computation ISO	Rank
Direct Instruction*	16	+065	2	-065	3	+656	1	+562	1
Southwest Lab	8	-250	5.5	000	1	+250	2	+062	3
Parent Education	12	+125	1	-042	2	-167	5	+042	4
Behavior Analysis	13	-038	3	-269	4	-115	4	+385	2
Bank Street	12	-208	4	-500	8.5	000	3	-125	7
TEEM (Arizona)	15	-367	7	-500	8.5	-200	7	+033	5
Responsive Education	16	-250	5.5	-406	7	-188	6	-062	6
Open Education (EDC)	10	-300	6	-400	6	-250	8	-300	9
Cognitive Curriculum	9	-389	8	-333	5	-500	9	-222	8
*Direct Instruction less Grand Rapids	14	+071		-035		+786		+643	

Table 2 gives a breakdown of the Basic Skills ISOs by variable using the All Data analysis. This breakdown shows that the positive ISOs for Direct Instruction were largely achieved through highly successful relative outcomes for MAT Language (+656) and MAT Math Computation (+562). Without Grand Rapids, these ISOs are +786 and +643, respectively. No model program produced a positive outcome on Spelling. Direct Instruction ranked second on Word Knowledge, third on Spelling, and first on Language and Math Computation.

Cognitive-Conceptual Skills

Figure 2 and Table 4 present the results for Cognitive-Conceptual Skills. Many expected the Direct Instruction Model to do well on Basic Skills, but not on Cognitive-Conceptual Skills. The data show this assumption to be incorrect. The Direct Instruction Model ranks first on Reading (comprehension), first on Math Concepts, and first on Math Problem Solving. The model does not perform particularly well on the Raven's (with relative comparisons), but neither does anyone else.

Figure 2, Table 4

More positive than negative results for Cognitive Measures are found for only one other model, Southwest Labs' Language Development Approach (SEDL). Substantially negative average ISOs are found for TEEM (-160), Behavior Analysis (-190), Open Education (-236) and Cognitive Curriculum (-249). Those programs whose espoused goals were to improve cognitive-conceptual skills were apparently unsuccessful. The relatively poor performance of Behavior Analysis on Cognitive Measures in comparison

FIGURE 2

ISO's for Cognitive Measures

(Reading Comprehension, Math Concepts, Math Problem Solving)

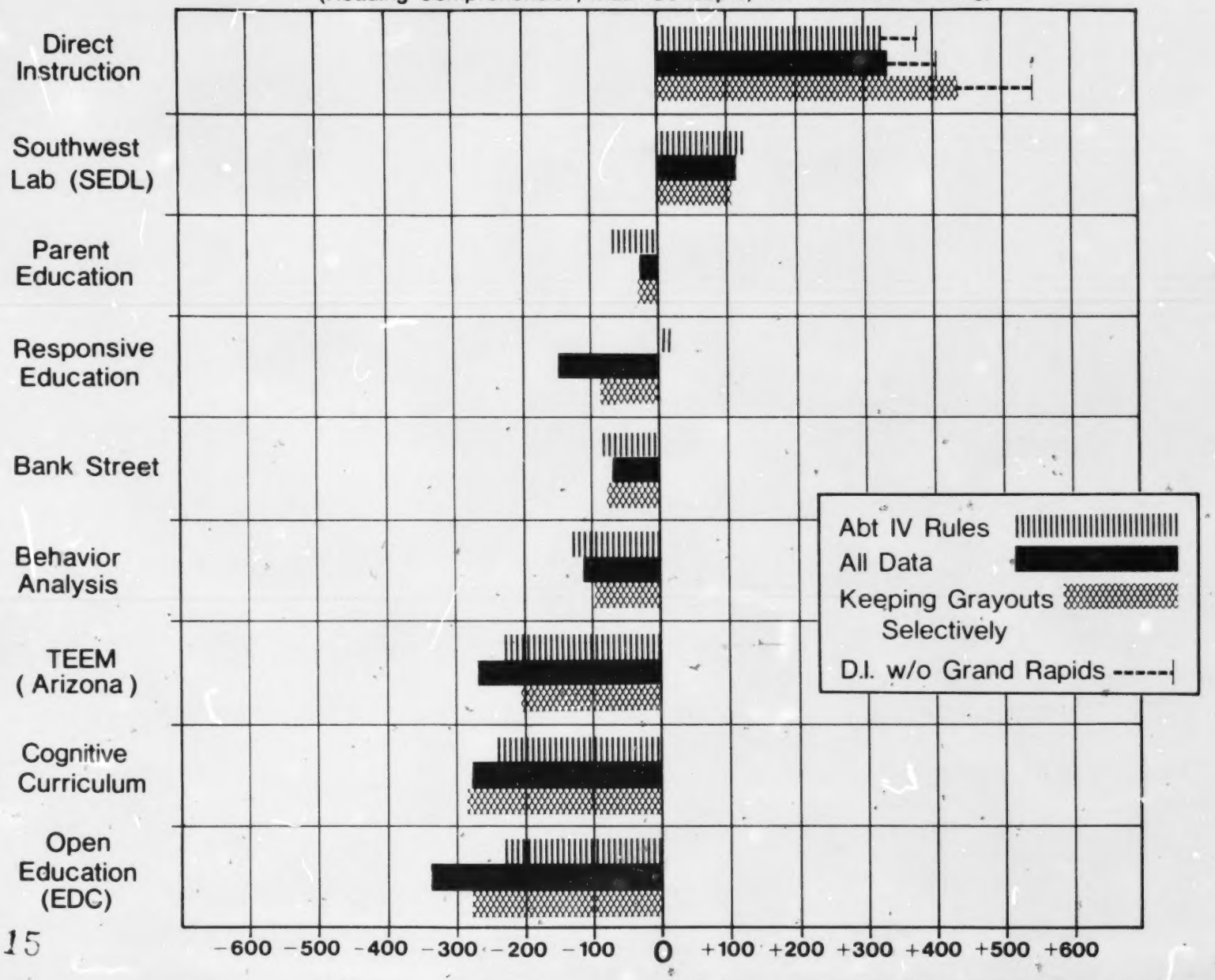


Table 4

ISOs for Cognitive Skill Subclasses Using All Data

	No. Sites	Reading ISO	Reading Rank	Math Concepts ISO	Math Concepts Rank	Math Problem Solving ISO	Math Problem Solving Rank	Raven's Progressive Matrices ISO	Raven's Progressive Matrices Rank
Direct Instruction*	16	+219	1	+312	1	+479	1	-200	7
Southwest Lab	8	+062	2	+250	2	+062	2	-062	4
Parent Education	12	000	4	-083	4	000	3	+042	2
Responsive Education	16	-094	5	-188	6	-156	6	-094	5
Bank Street	12	+042	3	-167	5	-083	4	-042	3
TEEM (Arizona)	15	-300	8	-333	8	-167	7	+067	1
Behavior Analysis	13	-154	6	-077	3	-115	5	-385	9
Open Education (EDC)	10	-400	9	-300	7	-300	9	-100	6
Cognitive Curriculum	9	-222	7	-389	9	-222	8	-222	8
*Direct Instruction less Grand Rapids	14	+214		+429		+571		-179	

to Basic Skills merits attention. It suggests that the systematic use of reinforcement procedures may not be enough in overcoming the current inability of schools to teach language competencies to students from economically disadvantaged homes.

In evaluating the capabilities of models to produce meaningful educational outcomes in the cognitive area, we believe the analyses without the Raven's and without Grand Rapids most clearly show what the Direct Instruction model can do. ISOs of from +375 to +544 suggest that even with some implementation problems, the model can be substantially effective in a wide variety of circumstances. The findings in the area of mathematical reasoning are especially impressive.

Affective Measures

Figure 3 and Table 5 present the ISOs for the Affective Measures. In the Abt IV Report these findings are discussed as follows (Abt Associates Inc., 1977, Vol. IV-B, p. 73):

> To some observers the performance of FT children in Direct Instruction sites on the affective measures is an unexpected result. The Direct Instruction model does not explicitly emphasize affective outcomes of instruction, but the sponsor has asserted that they will be the consequence of effective teaching. Critics of the model have predicted that the emphasis of the model on tightly controlled instruction might discourage children from freely expressing themselves and thus inhibit the development of self-esteem and other affective skills. In fact, this is not the case.

The findings again show the Direct Instruction Model to rank first on the Affective Measures. The different approaches to data analysis make little difference on these measures. It is interesting that the other

major sponsor focusing the use of positive reinforcement (Behavior Analysis) ranks second on these measures.

Figure 3, Table 5

Table 6 gives the rankings for Basic, Cognitive, and Affective Measures using ISOs based on the average of the three analytic methods. There is a clear correlation between positive academic outcomes and positive affective outcomes. The table also shows a substantial correlation between Basic and Cognitive Measures, which we believe should be expected. The major deviations in rankings are the relatively poor performance of Behavior Analysis on Cognitive Measures noted earlier, and the better ranking (although still negative ISOs) for Weikart's Cognitive Curriculum on the Affective Measures.

Table 6 here

In summary, these control-group comparisons clearly show the Direct Instruction Model to be effective in comparison to traditional programs (the comparison groups) and in comparison to the other model approaches to compensatory education having the same resources available. The consistency of these findings cannot easily be attributed to accident or fancy. When normative levels of performance are examined, this conclusion becomes even stronger.

Normative Performance

The Abt III and IV Reports provide median grade-equivalent scores by site and by sponsor for four MAT measures: Total Reading, Total Math, Spelling, and Language. The means for these data by model (converted to

FIGURE 3

ISO's for Affective Measures

(Coopersmith Self-Esteem, Intellectual Achievement Responsibility Scale)

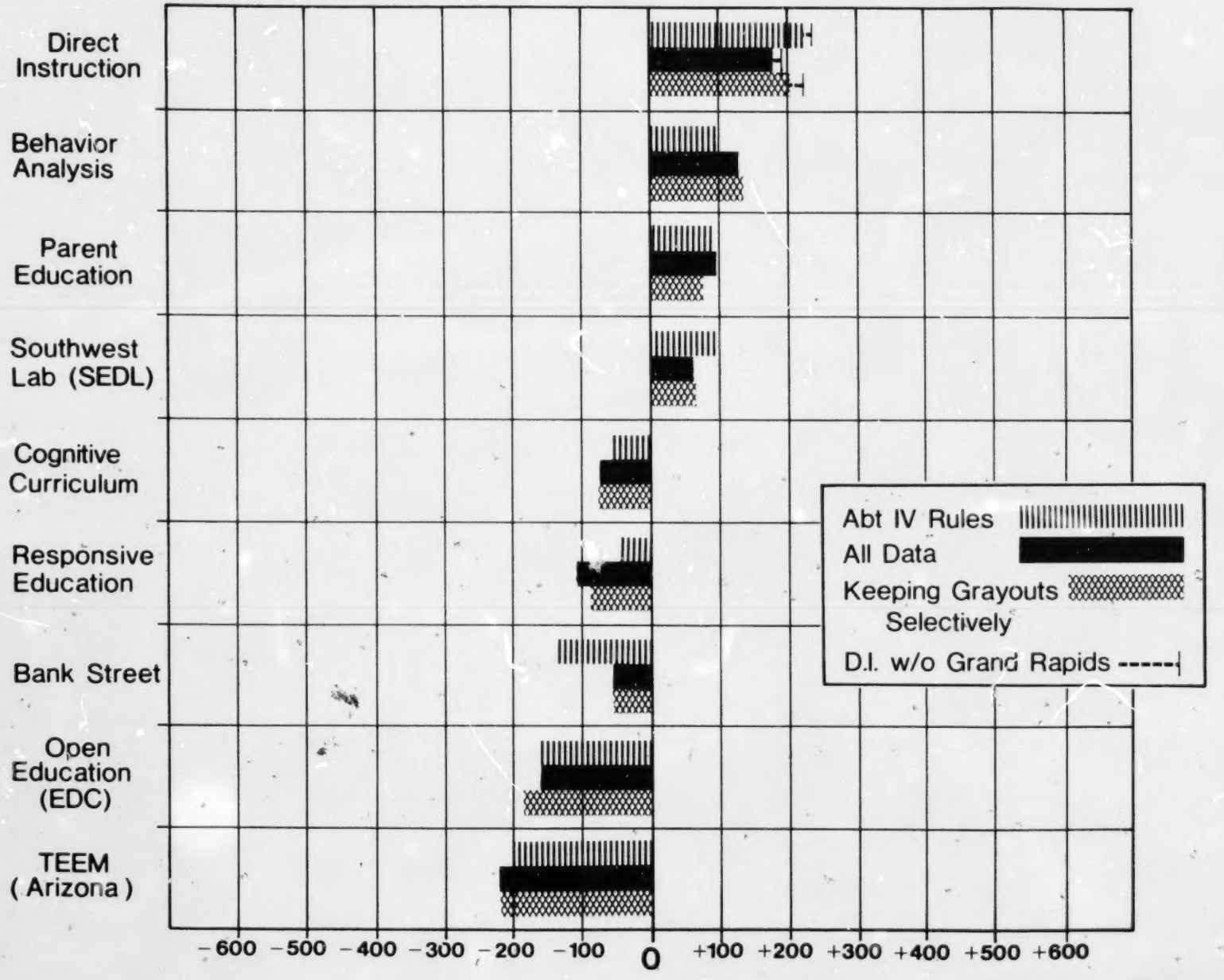


Table 5

ISOs for Affective Subclasses Using All data

	No. Sites	Coopersmith Self Esteem		IARS-		IARS+	
		ISO	Rank	ISO	Rank	ISO	Rank
Direct Instruction*	16	+156	3	+188	2	+188	1
Behavior Analysis	13	+231	1	+231	1	-077	5
Parent Education	12	+042	4.5	+125	3	+125	2
Southwest Lab	8	+188	2	000	4	000	3
Cognitive Curriculum	9	-056	8	-111	5	-056	4
Responsive Education	16	-031	7	-188	7	-094	7
Bank Street	12	+042	4.5	-125	6	-083	6
Open Education (EDC)	10	000	6	-250	9	-250	8
TEEM (Arizona)	15	-167	9	-233	8	-267	9
*Direct Instruction less Grand Rapids	14	+179		+179		+214	

Table 6

Ranks for Average ISOs by Area

	Basic Skills	Cognitive Measures (less Ravens)	Affective Measures	Average Rank
Direct Instruction	1	1	1	1
Southwest Lab	2	2	4	2
Parent Education	3	3	3	3
Behavior Analysis	4	6	2	4
Bank Street	5	4	7	5
Responsive Education	7	5	6	6
TEEM (Arizona)	6	7	9	7.5
Cognitive Curriculum	9	8	5	7.5
Open Education (EDC)	8	9	8	9

percentiles) are presented in Tables 7 through 10. The tables display percentiles on a one-fourth standard deviation scale. With this display, differences between sponsors of one-quarter standard deviation or more are easily detected and a norm reference is provided. The 20th percentile, which represents the average expectation for disadvantaged children without special help, was chosen for a baseline in drawing the graphs in Tables 7 to 10.[6]

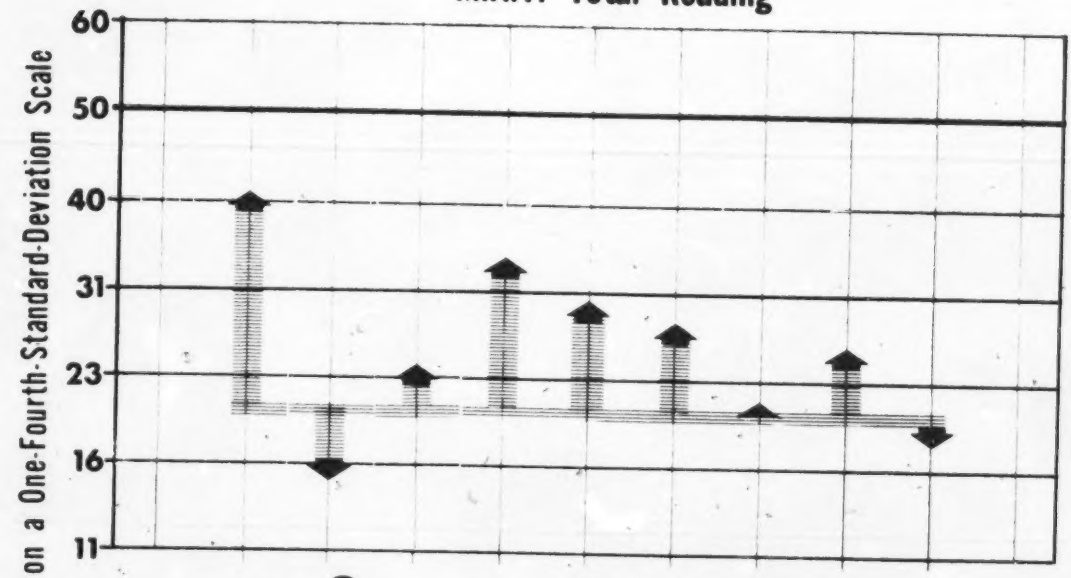
Tables 7, 8, 9, & 10

The Direct Instruction Model is clearly highest in normative performance and is close to or at national norms on all measures. These data include Grand Rapids. Without Grand Rapids the percentiles are 43 for Total Reading, 56 for Total Math, 54 for Spelling, and 56 for Language. Since models are ordered by average ISO rank (Table 6), a comparison of statistical and normative performance is readily made. The Direct Instruction program remains first in each case. SEDL and Parent Education models show the greatest drop in standing in going from ISOs to normative comparisons.

Only four sponsors have reading programs that are making some headway toward average performance by the end of third grade (Direct Instruction, Behavior Analysis, Bank Street College, and Responsive Education). In Math, Direct Instruction is at least one-half standard deviation ahead of all the others. In Spelling, the Behavior Analysis program is the only program other than Direct Instruction approaching national norms. In Language (usage, punctuation, and sentence types) the Direct Instruction program is three-fourths of a standard deviation ahead of all other programs.

TABLE 7

M.A.T. Total Reading

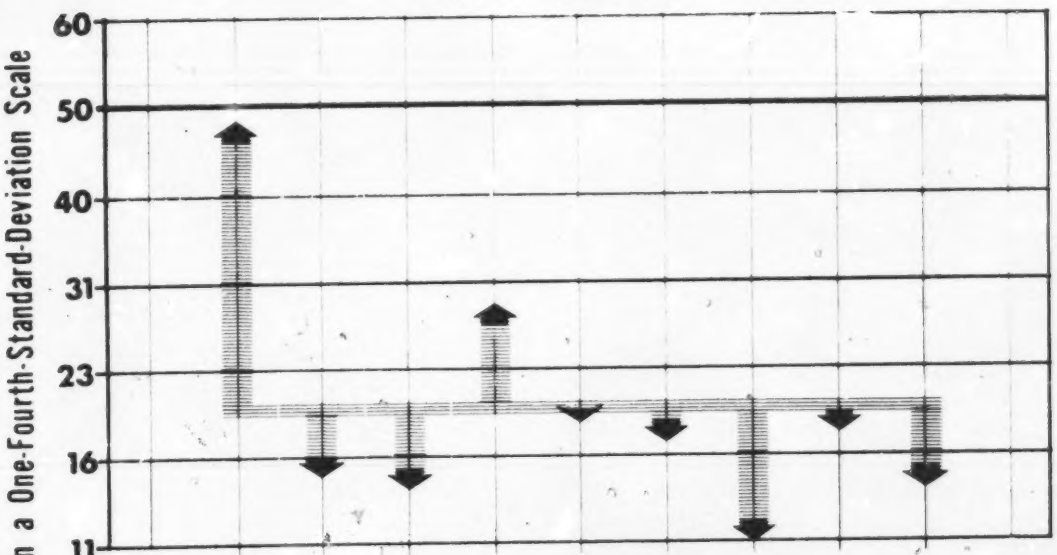


	Direct Instruction	Southwest Lab (SEDL)	Parent Education	Behavior Analysis	Bank Street	Responsive Education	Cognitive Curriculum	TEEM (Arizona)	Open Education (EDC)
%-tile	41	15	24	34	30	28	21	26	18
S. S.	58.3	49.6	52.8	55.8	55.1	53.8	51.5	53.7	50.6
G. E.	3.23	2.43	2.68	2.98	2.91	2.78	2.55	2.77	2.48

Grade equivalent for 50th percentile is 3.5

TABLE 8

M.A.T. Total Math

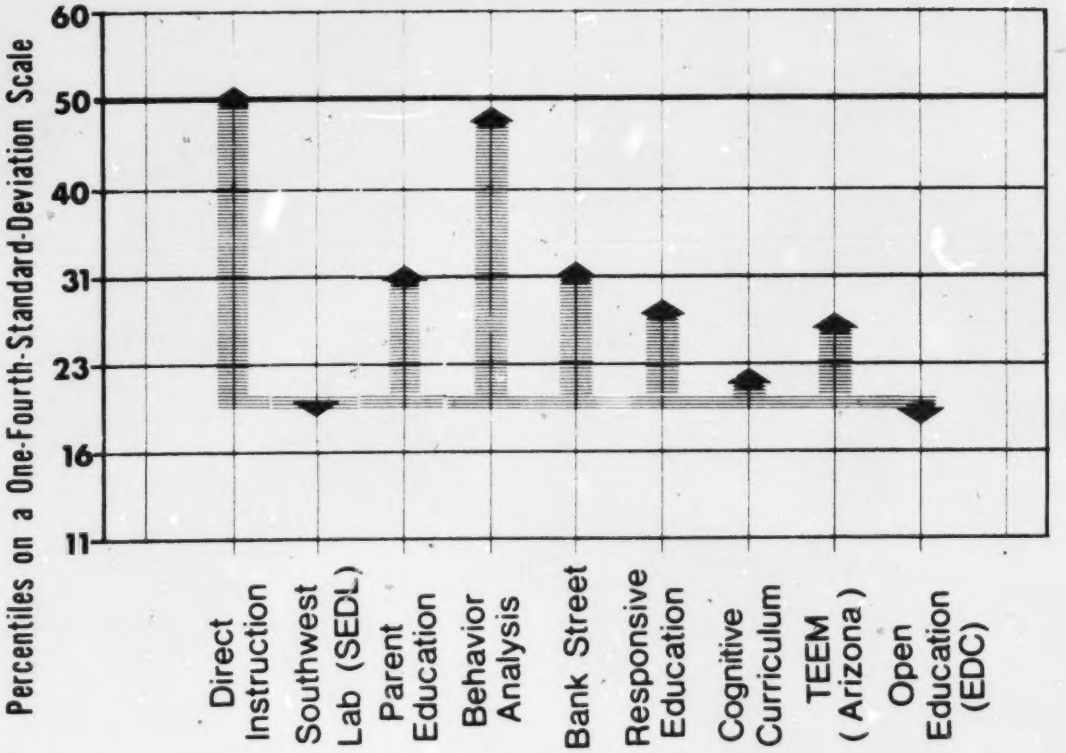


	Direct Instruction	Southwest Lab (SEDL)	Parent Education	Behavior Analysis	Bank Street	Responsive Education	Cognitive Curriculum	TEEM (Arizona)	Open Education (EDC)
%-tile	48	15	14	28	19	17	11	18	14
S. S.	70.1	58.5	57.8	63.8	60.6	59.3	55.8	60.0	58.1
G. E.	3.71	2.85	2.78	3.24	3.03	2.93	2.58	3.00	2.81

Grade equivalent for 50th percentile is 3.75

TABLE 9

M.A.T. Spelling

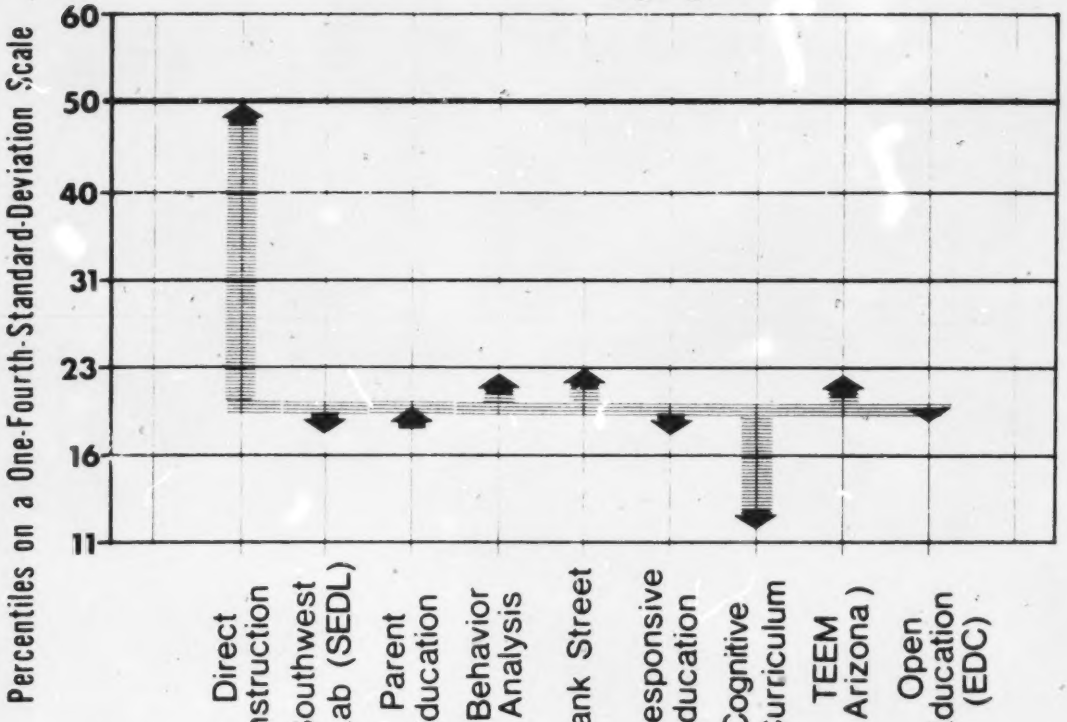


	Direct Instruction	Southwest Lab (SEDL)	Parent Education	Behavior Analysis	Bank Street	Responsive Education	Cognitive Curriculum	TEEM (Arizona)	Open Education (EDC)
%-tile	51	19	32	49	32	28	22	27	18
S. S.	64.3	52.6	58.0	63.6	58.0	56.6	54.1	56.4	51.8
G. E.	3.63	2.53	2.95	3.56	2.95	2.86	2.61	2.84	2.48

Grade equivalent for 50th percentile is 3.6

TABLE 10

M.A.T. Language



	Direct Instruction	Southwest Lab (SEDL)	Parent Education	Behavior Analysis	Bank Street	Responsive Education	Cognitive Curriculum	TEEM (Arizona)	Open Education (EDC)
%-tile	50	18	20	22	23	18	12	22	19
S. S.	70.0	56.0	57.2	58.0	58.6	56.1	54.1	58.5	56.6
G. E.	4.00	2.60	2.72	2.80	2.86	2.61	2.41	2.85	2.66

Grade equivalent for 50th percentile is 4.0

The normative performance data add much to the interpretation of the statistical comparisons. The Direct Instruction program is apparently able to make effective use of teaching time in most areas of skill development so as to show a superior outcome nearly across the board.[7]

Site Variability

A major conclusion in the Abt IV Report supported in the House critique is this:

"Any model can 'fail' by having a group perform lower on a test than it would have without Follow Through" (Abt Associates Inc., 1977, Vol. IVA, p. 140). With respect to the Direct Instruction Model, this statement fails to recognize that we did not serve as sponsors for Grand Rapids for nearly two years for the Cohort II children, and three years for the Cohort III children. But the Abt Report and the House critique also err in the logical basis of their conclusion. The conclusion about variability is based on control-group comparisons. The sponsor did not select the control groups nor monitor any special programs they might have (and many did have special programs through other Federal and local funds). In some of our sites, our model performs at or near grade level for specific measures and yet a minus (-) occurs in the summary table. This kind of variability should not be attributed to a model weakness. If all disadvantaged groups performed at grade level, there would be no need for Follow Through. The question of implementation variability is more appropriately explored with norm-referenced measures of "absolute performance" rather than indices of significance. The Direct Instruction Model median grade equivalents on the MAT for K-starting sites is given in Table 11.

Insert Table 11

Table 11

Direct Instruction Model Site Variation in

Median Performance for K-Starting Sites

Total Reading	Total Math	Spelling	Language
3.7	4.2	4,3	4.8
Norm (3.5)→3.5	4.0	4,3	4.8*
3.4	4.0	3,9	4.7*
3.4	3.9	3,9	4.6*
3.3	3.9	3.7	4.5
3.2	3.8	Norm (3.6)→	4.2*
3.1	3.8	3.4	4.1
3.1	Norm (3.75)→	3.4	Norm (4.0)→ 4.0
(2.8)	3.7	3.4	(2.8)*
(2.8)	(3.0)	(3.0)	(2.6)
	(2.8)	(3.0)	

( ) Grand Rapids

Norm = 50th percentile

*Means are substituted for missing medians

Without Grand Rapids, the norm data for Math, Spelling, and Language show that most of the variability is above the grade norm. In Reading this is not the case, but the range below grade norm is only four tenths of a grade level. As we have argued elsewhere MAT Reading at the Elementary Test level is heavily dependent on middle class home language inputs rather than what is taught in schools (Becker, 1977) and consequently, the reading skills are more difficult to impact than math and spelling. The point is that except for reading, the variability indicates that school means for economically disadvantaged children are near or above the national norm. Variability of this type is hardly a problem.

Summary

A simple examination of the Abt significance data and the Abt normative data by sponsors shows clearly that Direct Instruction was more effective than any of the other Follow Through Models, and the degree of effectiveness was non-trivial. Bereiter and Kurland (1978) draw a similar conclusion from their very sophisticated reanalysis of the basic Abt data to show the significance of differences between various sponsors.

Footnotes

[1]Parts of this paper are summarized from the author's book in preparation (Engelmann & Becker, Forward from Basics--The Case for Direct Instruction). All rights are reserved by the authors.

The University of Oregon Follow Through Project was supported by funds from the U.S. Office of Education, Department of School Systems, Division of Compensatory Education, Follow Through Branch, under Grants OEG-070-4257 (286) and G00-7507234.

The authors are indebted to their many coworkers in this endeavor - office staff, data staff, project managers, project directors, supervisors, teachers, aides, parents, and children who made this work possible.

[2]In the entering-kirdergarten streams, 25% of the differences called null were educationally significant but not statistically significant, probably because of attrition effects on sample size. Only 5% were statistically significant but not educationally significant (Abt Associates, 1977, p. 132).

[3]The grayout because of preschool differences eliminates effects as follows:

	Kind of Effect			
	+	0	-	Total
Direct Instruction	22	35	3	60
Eight other Major Sponsors	16	262	127	405

In percentages, the preschool based grayouts eliminate 37% positive outcomes and 5% negative outcomes for Direct Instruction, while the same procedure eliminates 4% positive outcomes and 31% negative outcomes for the other major sponsors. Two models, Cognitive Curriculum and Responsive Education, especially benefit by this inappropriate elimination of data. Responsive Education has 51 negatives and 1 positive eliminated (24% of all their outcomes), while

Cognitive Curriculum has 41 negatives and 5 positives eliminated (the negatives are 34% of all their outcomes). In other words, the effects of this questionable procedure for eliminating data is to greatly attenuate the differences between successful and unsuccessful models.

[4]In another publication (Becker & Engelmann, in press) we have published data summaries using ISOs which differ from those presented here in that they included all data grayed out by Abt Associates, and used the following rule for counting pluses and minuses: Only one of the analyses (local or pooled) had to be plus (or minus) to count the effect plus (or minus) as long as the other was not minus (or plus). In the two cases where one analysis was plus and one was minus, the effect was counted zero. The effect of this approach was to increase the magnitude of the negative outcomes for some sponsors, but the basic relations among sponsors on the ISOs was nearly the same. The Responsive Education model improved the most using the present analytic rules.

[5]In Grand Rapids, Michigan, a change of program directors in 1971 led to the site failing to continue implementation of the Direct Instruction Model. In the middle of the 1972 school year, after struggling to get the site to follow model procedures, we withdrew as sponsor. An attempt to re-establish the model in 1973 failed. Grand Rapids continued to be funded and continued in the National Evaluation as a Direct Instruction site, even though neither of these actions was justified in terms of the Follow Through agreements.

[6]A U.S. Office of Education report (1976) substantiates this approach: "Analyses of all test scores showed that the typical student who received compensatory assistance in reading was at the 20th percentile for grade 2

and the 22nd percentile for grades 4 and 6" (p. 88). Moreover, in a footnote to page 88 this additional information is given: "In conjunction with the Emergency School Aid Act evaluation, children in grades 3, 4, and 5 of a nationally representative sample of minority isolated schools (50% or more non-white) perfo ed at the 23rd, 18th and 19th percentiles, respectively, on reading achievement in the Spring of 1973; similar results were obtained for mathematics achievement" (Ozenne, D. G., et al., 1974). The educational requirement for Title I eligibility (one year or more below grade level) is the 20th percentile for Metropolitan Total Reading.

[7]Sponsor collected data from 13 sites and 4 to 6 cohorts show normative results nearly identical to those from the Abt report (Becker & Engelmann, 1976, 1978).

References

Abt Associates. Education as experimentation: A planned variation model (Vol. III). Cambridge, Mass.: Abt Associates, 1976.

Abt Associates. Education as experimentation: A planned variation model (Vol. IV). Cambridge, Mass.: Abt Associates, 1977.

Becker, W. C. Critique of Abt IV - Follow Through Final Report. (Mimeo.) University of Oregon, College of Education, Eugene, OR, 1977.

Becker, W. C. Teaching reading and language to the disadvantaged--What we have learned from field research. Harvard Education Review, 1977, 47, 518-543.

Becker, W. C., & Engelmann, S. Analysis of achievement data on six cohorts of low income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Follow Through Project, Technical Report #76-1. Unpublished manuscript. University of Oregon, 1976.

Becker, W. C., & Engelmann, S. Analysis of achievement data on six cohorts of low income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Follow Through Project, Technical Report #78-1. Unpublished manuscript. University of Oregon, 1978.

Becker, W. C., & Engelmann, S. The direct instruction model. In R. Rhine (Ed.), Encouraging change in America's schools: A decade of experimentation. New York: Academic Press, in press.

Bereiter, C. and Kurland, M. Were some Follow Through Models more effective than others? Paper presented at the American Educational Research Association meeting, Toronto, Canada, March 1978.

Campbell, D. T. and Erlebacher, A. How regression artifacts in quasi-experimental evaluations in compensatory education tend to underestimate effects. In J. Hellmuth (Ed.), Disadvantaged Child. Compensatory Education: A National Debate (Vol. 3). New York: Brunner/Mazel, 1970, pp. 185-225.

Coopersmith, S. The antecedents of self-esteem. San Francisco: W. H. Freeman, 1967.

Crandall, V. C., Katkowsky, W. & Crandall, V. J. Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. Child Development, 1965, 36, 91-109.

Engelmann, S., & Becker, W. C. Forward from basics--the case for direct instruction. In preparation.

House, E. R., Glass, G. V., McLean, L. D., Walker, D. E., No simple answer: critique of the "Follow Through" evaluation: mimeo report. Urbana, IL, University of Illinois, 1977.

House, E. R., Glass, G. V., McLean, L. D., Walker, D. E., Critique of the "Follow Through" evaluation. Phi Delta Kappan, Vol. 59, No. 7, March, 1978, pp. 473-74.

McDaniels, G. L. Evaluation of problems in Follow Through. In Rivlin, A. M., & Timpane, P. M. (Eds.) Planned variation in education. Washington, D. C.: The Brookings Institution, 1975.

United States Office of Education. Annual evaluation report on programs administered by the U.S. Office of Education, FY 1975. Washington, D.C.: Capital Publications, Educational Resources Division, 1976.